

Locally adaptive Bayesian covariance regression

Daniele Durante¹, Bruno Scarpa^{1,*}, David B. Dunson²

¹Department of Statistical Sciences, University of Padua, Padua, Italy

²Department of Statistical Sciences, Duke University, Durham, NC, USA

**email*: scarpa@stat.unipd.it

Abstract

Multivariate time series data arise in many applied domains, and it is often crucial to obtain a good characterization of how the covariance among the different variables changes over time. Certainly this is the case in financial applications in which covariance can change dramatically during times of financial crisis, revealing different associations among assets and countries than occur in a healthier economic climate. Our focus is on developing models that allow the covariance to vary flexibly over continuous time, and additionally accommodate locally adaptive smoothing of the covariance. Locally adaptive smoothing to accommodate varying smoothness in a trajectory over time has been well studied, but such approaches have not yet been developed for time-varying covariance matrices to our knowledge. To address this gap, we generalize recently develop methods for Bayesian covariance regression to incorporate random dictionary elements with locally varying smoothness. Using a differential equation representation, we additionally develop a fast computational approach via MCMC, with online algorithms also considered. The performance of the models is assessed through simulation studies and the methods are applied to financial time series.

Key Words: Bayesian nonparametrics; locally varying smoothness; long-range dependence; multivariate time series; nested Gaussian process; stochastic volatility.

1. Introduction

In the field of multivariate time series, modeling the time-varying covariance structure in high dimensional datasets is a key issue in many application domains. For example, in our motivating application, the focus is on capturing volatility and co-volatilities of different assets and countries. There are many other applications that have arisen with the routine collection of online data, such as Google Trends. When the dynamics of the series are

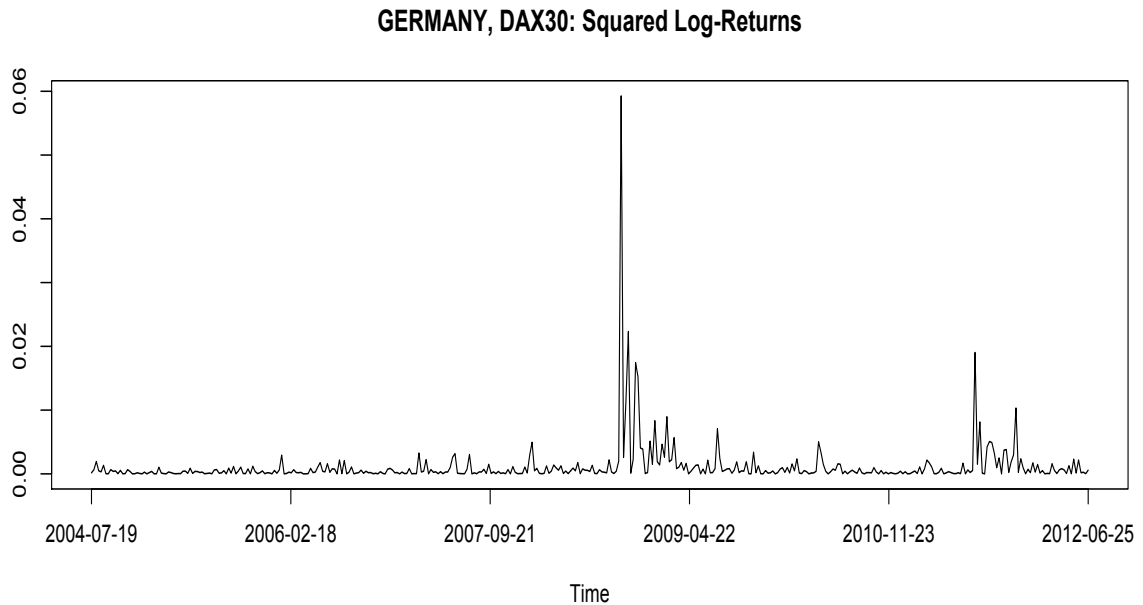


Figure 1: Squared Log>Returns of DAX30, using weekly data from 2004/07/19, to 2012/06/25.

characterized by dramatic changes, it is important to incorporate locally adaptive smoothing of not only the mean but also the covariance. Figure 1 shows an example from financial time series, where volatilities, described by the squared log-returns of the German stock market index (DAX30), change dramatically, motivating the search for flexible models able to capture such rapid changes.

There is a rich literature on univariate stochastic volatility modeling, with an increasing emphasis on multivariate generalizations. One popular approach estimates the $p \times p$ time-varying covariance matrix Σ_t via an exponentially weighted moving average (EWMA; see, e.g., Tsay, 2005). This approach uses a single time-constant smoothing parameter $0 < \lambda < 1$, with extensions to accommodate locally-varying smoothness λ_t not straightforward due to the need to maintain positive semidefinite Σ_t at every time. To address this problem, generalizations of EWMA have been proposed including the diagonal vector ARCH model (DVEC), (Bollerslev, Engle and Wooldridge, 1988) and its variant, the BEKK model (Engle and Kroner, 1995). However, these models are computationally demanding and are not designed for moderate to large p . Principal component GARCH (Ding, 1994) and O-GARCH (Alexander, 2001) perform dimensionality reduction through a latent factor model (see also van der Weide, 2002). However, time-constant factor loadings and uncorrelated latent factors

restrict evolution of Σ_t .

Such models fall far short of our goal of allowing Σ_t to be fully flexible with the dependence between Σ_t and $\Sigma_{t+\Delta}$ varying with not just the time-lag Δ but also time. In addition, these models do not handle missing data easily and tend to require long series for accurate estimation (Burns, 2005). Accommodating changes in continuous time are also important in many applications and to avoid having the model be critically dependent on the time scale, with inconsistent models obtained as time units are varied.

An alternative to multivariate GARCH is to characterize the volatility process as a random function of past returns (Harvey, 1994). Missing values can be handled but most of the literature limits flexibility in evolution of Σ_t by using models that characterize time changes solely through variances. To address this problem, Philipov and Glickman (2006a) develop a multivariate stochastic volatility model in which the time-varying covariance structure follows a stochastic process based on the Wishart distribution. Computational challenges arise in implementations, and the use of the Wishart limits generalizations of moderate to large p settings. Prado and West (2010) address the problem of posterior computation for dynamic covariance matrices via discounting methods that maintain simple update equations as new observations are added. However, the model restricts the evolution of the covariance to be stationary and slowly-changing.

Even for iid observations from a multivariate normal model with a single time stationary covariance matrix, there are well known problems with Wishart priors motivating a rich literature on dimensionality reduction techniques based on factor and graphical models. There has been abundant recent interest in applying such approaches to dynamic settings. Refer to Nakajima and West (2012) and the references cited therein for recent literature on Bayesian dynamic factor models for multivariate stochastic volatility. Their approach allows the factor loadings to evolve dynamically over time, while including sparsity through a latent thresholding approach, leading to apparently improved performance in portfolio allocation. They utilize a time-varying discrete-time autoregressive model, which allows the dependence in the covariance matrices Σ_t and $\Sigma_{t+\Delta}$ to vary as a function of both t and Δ . However, the result is an extremely richly parameterized and computationally challenging model, with selection of the number of factors proceeding by cross validation. Our emphasis is instead on developing continuous time stochastic processes for Σ_t , which accommodate locally-varying smoothness.

In this regard, a highly relevant development was the Bayesian covariance regression (BCR) model of Fox and Dunson (2011), which defines the covariance matrix as a regularized quadratic function of time-varying loadings in a latent factor model, characterizing the latter as a sparse combination of a collection of unknown Gaussian process (GP) dictionary functions. Although their approach provides a continuous time and highly flexible model that accommodates missing data and scales to large p , there are two limitations motivating this article. Firstly, their proposed covariance stochastic process assumes a stationary dependence structure, and hence tends to under-smooth during periods of stability and over-smooth during periods of dramatic change. Secondly, the well known computational problems with usual GP regression are inherited, leading to difficulties in scaling to long series and issues in mixing of MCMC algorithms for posterior computation.

In this article, we address both of these problems to develop a novel covariance stochastic process with locally-varying smoothness. This is accomplished by modifying the method of Fox and Dunson (2011) to incorporate dictionary functions that are assigned nested Gaussian process (nGP) priors (Zhu and Dunson, 2012). Such nGP priors reduce the GP computational burden involving matrix inversions from $O(T^3)$ to $O(T)$, with T denoting the length of the time series, while also allowing flexible locally-varying smoothness. Marginalizing out the dictionary functions, we obtain a covariance stochastic process that inherits these advantages. We also propose an online implementation, which can accommodate streaming data.

In Section 2, we describe the basic model structure with particular attention to prior specification. Section 3 explores the main features of the Gibbs sampler for posterior computation and outlines the steps for a fast online updating approach. In Section 4 we compare our model to BCR through a simulation study. Finally in Section 5 an application to stock market indices across countries is examined.

2. Covariance Regression with Locally Adaptive Smoothing Priors

2.1 Notation and Motivation

Let y_i represent a $p \times 1$ vector of observations where $i = 1, \dots, T$ indexes time with

$$y_i \sim N_p(\mu(t_i), \Sigma(t_i)),$$

with $\mu(t_i)$ and $\Sigma(t_i)$ denoting, respectively, the $p \times 1$ mean vector and $p \times p$ covariance

matrix at time $t_i \in \mathcal{T} \subset \mathbb{R}^+$, where t_i represents the time of the i th observation and observations can be unequally-spaced. Following a Bayesian approach, we define priors for $\Sigma_{\mathcal{T}} = \{\Sigma(t), t \in \mathcal{T}\}$ and $\mu_{\mathcal{T}} = \{\mu(t), t \in \mathcal{T}\}$, which are designed to be flexible, accommodate locally-varying smoothness in time, be defined in continuous time, and facilitate scaling both computationally and statistically to moderately large p .

2.2 Latent Factor Model

A common strategy in modeling of large covariance matrices is to rely on a lower-dimensional factorization, with factor analysis providing one possible direction. Sparse Bayesian factor models have been particularly successful in challenging cases, while having advantages over frequentist competitors in incorporating a probabilistic characterization of uncertainty in the number of factors as well as the parameters in the loadings and residual covariance. For recent articles on Bayesian sparse factor analysis for a single large covariance matrix, refer to Bhattacharya and Dunson (2011), Pati et al. (2012) and the references cited there-in.

In our setting, we are instead interested in letting the covariance matrix vary flexibly over time. Extending the usual factor analysis framework to this setting, we let

$$\Sigma(t) = \Lambda(t)\Lambda(t)^T + \Sigma_0$$

where $\Lambda(t)$ is a $p \times K$ time-varying factor loadings matrix and $\Sigma_0 = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. There is a literature on using Bayesian factor analysis with time-varying loadings, but essentially all the literature assumes discrete-time dynamics on the loadings while our focus is instead of allowing the loadings, and hence the induced covariance $\Sigma(t)$, to evolve flexibly in continuous time. Hence, we are most closely related to the literature on Gaussian process latent factor models for spatial and temporal data; refer, for example, to Lopes, Salazar and Gamerman (2008) and Lopes, Gamerman and Salazar (2011). In these models, the factor loadings matrix characterizes spatial dependence, with time varying factors accounting for dynamic changes. Fox and Dunson (2011) instead allow the loadings matrix to vary through a continuous time stochastic process built from latent GP dictionary functions. The basic structure of our model is identical to theirs, but we propose a fundamentally different prior on the dictionary elements and approach to computation.

Model (1) can be induced by marginalizing out the latent factors η_i in

$$y_i = \Lambda(t_i)\eta_i + \epsilon_i \tag{1}$$

where $\eta_i \sim N_K(0, I_K)$ and $\epsilon_i \sim N_p(0, \Sigma_0)$. We assume that the time-varying factor loadings matrix $\Lambda(t)$ is a linear combination of a much smaller set of continuous dictionary functions $\xi_{lk} : \mathcal{T} \rightarrow \mathfrak{R}$ comprising the $L \times K$, with $L \ll p$, matrix $\xi(t)$. As a result

$$\Lambda(t_i) = \Theta \xi(t_i)$$

where Θ is a $p \times L$ matrix of coefficients relating the matrix $\Lambda(t)$ to the time-varying dictionary elements in $\xi(t)$. Such a decomposition further reduces the number of continuous random functions to be modeled from $p \times K$ to $L \times K$ leading to a more computationally tractable formulation in which the induced covariance structure, after marginalizing out the latent factors, follows the equation

$$\text{cov}(y_i | t_i = t) = \Sigma(t) = \Theta \xi(t) \xi(t)^T \Theta^T + \Sigma_0 \quad (2)$$

The above decomposition of $\Sigma(t)$ is not unique. Potentially we could constrain the loadings matrix to enforce identifiability (Geweke and Zhou, 1996), but this approach induces an undesirable order dependence among the responses (Aguilar and West, 2000, West, 2003, Lopes and West, 2004, Carvalho et al., 2008). Given our focus on estimation of $\Sigma(t)$, we follow Ghosh and Dunson (2009) in avoiding identifiability constraints, as such constraints are not necessary to ensure identifiability of the induced covariance $\Sigma(t)$. The characterization of the class of time-varying covariance matrices $\Sigma(t)$ is proved by Lemma 2.1 of Fox and Dunson (2011) which states that for K and L sufficiently large, any covariance regression can be decomposed as in (2).

When $\mu(t)$ is not know, we can incorporate nonparametric mean regression in the model formulation by letting

$$\eta_i = \psi(t_i) + \nu_i, \quad \nu_i \sim N_K(0, I_K), \quad (3)$$

where $\psi(t_i) = [\psi_1(t_i), \dots, \psi_K(t_i)]^T$ is a vector of continuous functions $\psi_j : \mathcal{T} \rightarrow \mathfrak{R}$ that can be modeled similarly to the dictionary functions ξ_{lk} . As a result, marginalizing out the latent factors, the induced mean of y_i conditionally on $t_i = t$ is

$$\mu(t) = \Theta \xi(t) \psi(t) \quad (4)$$

2.3 Prior Specification

We specify independent priors Π_ξ , Π_Θ , Π_{Σ_0} and Π_ψ respectively for $\xi_{\mathcal{T}} = \{\xi(t), t \in \mathcal{T}\}$, Θ ,

Σ_0 and $\psi_{\mathcal{T}} = \{\psi(t), t \in \mathcal{T}\}$ in (2) and (4), to induce priors Π_{Σ} and Π_{μ} on $\Sigma_{\mathcal{T}}$ and $\mu_{\mathcal{T}}$ with the goal of maintaining simple computation and allowing both covariances and means to vary flexibly over continuous time. Fox and Dunson (2011) address this issue by considering dictionary function as $\xi_{lk} \sim GP(0, c)$ independently for all l, k , with c the squared exponential correlation function having $c(\xi, \xi') = \exp(-\kappa \|\xi - \xi'\|_2^2)$. To allow locally-adaptive smoothing and facilitate computation, we induce a prior on the dictionary functions through stochastic differential equations.

Specifically, following the nested Gaussian process (nGP) formulation of Zhu and Dunson (2012), we let

$$D^m \xi_{lk}(t) = A_{lk}(t) + \sigma_{\xi_{lk}} W_{\xi_{lk}}(t), \quad m \in N, \quad m \geq 2, \quad (5)$$

$$D^n A_{lk}(t) = \sigma_{A_{lk}} W_{A_{lk}}(t), \quad n \in N, \quad n \geq 1, \quad (6)$$

where A_{lk} is the locally instantaneous mean, $\sigma_{\xi_{lk}} \in \mathbb{R}^+$, $\sigma_{A_{lk}} \in \mathbb{R}^+$, $W_{\xi_{lk}} : \mathcal{T} \rightarrow \mathbb{R}$ and $W_{A_{lk}} : \mathcal{T} \rightarrow \mathbb{R}$ are independent Gaussian white noise processes with mean $E[W_{\xi_{lk}}(t)] = E[W_{A_{lk}}(t)] = 0$, for all $t \in \mathcal{T}$, and covariance function $E[W_{\xi_{lk}}(t)W_{\xi_{lk}}(t')] = E[W_{A_{lk}}(t)W_{A_{lk}}(t')] = \delta(t - t')$ a delta function. This formulation naturally induces a prior for ξ_{lk} with varying smoothness, where $E[D^m \xi_{lk}(t) | A_{lk}(t)] = A_{lk}(t)$, and initialization at t_1 based on the assumption

$$[\xi_{lk}(t_1), D^1 \xi_{lk}(t_1), \dots, D^{m-1} \xi_{lk}(t_1)]^T \sim N_m(0, \sigma_{\mu_{lk}}^2 I_m).$$

The same goes for the initial value $A_{lk}(t_1)$ and its derivatives up to order $n - 1$ leading to the prior

$$[A_{lk}(t_1), D^1 A_{lk}(t_1), \dots, D^{n-1} A_{lk}(t_1)]^T \sim N_n(0, \sigma_{\alpha_{lk}}^2 I_n).$$

The prior for the initial values and for $W_{\xi_{lk}}$ and $W_{A_{lk}}$ are assumed mutually independent. Finally, we let

$$\begin{aligned} \sigma_{\xi_{lk}}^2 &\sim \text{invGa}(a_{\xi}, b_{\xi}) \\ \sigma_{A_{lk}}^2 &\sim \text{invGa}(a_A, b_A) \end{aligned}$$

independently for each (l, k) ; where $\text{invGa}(a, b)$ denotes the Inverse Gamma distribution with shape a and scale b .

The Markovian property implied by SDEs in (5) and (6) represents a key advantage in terms of computational tractability as it allows a simple state space formulation. In

particular, referring to Zhu and Dunson (2012) for $m = 2$ and $n = 1$ (this can be easily extended for higher m and n), and for $\delta_i = t_{i+1} - t_i$ sufficiently small, the prior for ξ_{lk} along with its first order derivative ξ'_{lk} and A_{lk} follow the approximated state equation

$$\begin{bmatrix} \xi_{lk}(t_{i+1}) \\ \xi'_{lk}(t_{i+1}) \\ A_{lk}(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_{lk}(t_i) \\ \xi'_{lk}(t_i) \\ A_{lk}(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i,\xi_{lk}} \\ \omega_{i,A_{lk}} \end{bmatrix}, \quad (7)$$

where $[\omega_{i,\xi_{lk}}, \omega_{i,A_{lk}}]^T \sim N_2(0, V_{i,lk})$, with $V_{i,lk} = \text{diag}(\sigma_{\xi_{lk}}^2 \delta_i, \sigma_{A_{lk}}^2 \delta_i)$. There are two crucial aspects to highlight. Firstly, this formulation allows continuous time and an irregular grid of observations over t by relating the latent states at $i + 1$ to those at i through the distance between t_{i+1} and t_i where i represents a discrete order index and $t_i \in \mathcal{T}$ the time observation related to the i th observation. Secondly, compared to Zhu and Dunson (2012) our approach represents an important generalization in: (i) extending the analysis to the multivariate case (i.e. y_i is a p -dimensional vector instead of a scalar) and (ii) accommodating locally adaptive smoothing not only on the mean but also on the time-varying variance and covariance functions.

To address the issue related to the selection of the number of dictionary elements a shrinkage prior Π_Θ is proposed for Θ . In particular, as proposed in Bhattacharya and Dunson (2011) we assume:

$$\begin{aligned} \theta_{jl} | \phi_{jl}, \tau_l &\sim N(0, \phi_{jl}^{-1} \tau_l^{-1}) \quad \phi_{jl} \sim Ga(3/2, 3/2) \\ \vartheta_1 &\sim Ga(a_1, 1), \quad \vartheta_h \sim Ga(a_2, 1), h \geq 2, \quad \tau_l = \prod_{h=1}^l \vartheta_h \end{aligned} \quad (8)$$

Note that if $a_2 > 1$ the expected value for ϑ_h is greater than 1. As a result, as l goes to infinity, τ_l tends towards infinity shrinking θ_{jl} towards zero. This leads to a flexible prior for θ_{jl} with a local shrinkage parameter ϕ_{jl} and a global column-wise shrinkage factor τ_l which allows many elements of Θ being close to zero as L increases.

Finally for the variances of the error terms in vector ϵ_i , we assume the usual inverse gamma prior distribution. Specifically Π_{Σ_0} is defined through

$$\sigma_j^{-2} \sim Ga(a_\sigma, b_\sigma)$$

independently for each $j = 1, \dots, p$. To conclude prior specification, if $\mu(t)$ is unknown, a similar approach based on dictionary elements can be considered in the definition of the

prior Π_ψ . In particular, recalling the previous results of the prior for ξ_{lk} , we can represent the prior for ψ_k with the following state equation

$$\begin{bmatrix} \psi_k(t_{i+1}) \\ \psi'_k(t_{i+1}) \\ B_k(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \psi_k(t_i) \\ \psi'_k(t_i) \\ B_k(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i,\psi_k} \\ \omega_{i,B_k} \end{bmatrix} \quad (9)$$

independently for $k = 1, \dots, K$, where $[\omega_{i,\psi_k}, \omega_{i,B_k}]^T \sim N_2(0, S_{i,k})$, with $S_{i,k} = \text{diag}(\sigma_{\psi_k}^2 \delta_i, \sigma_{B_k}^2 \delta_i)$. Similarly to ξ_{lk} , the priors for the initial values are assumed

$$\begin{aligned} [\psi_k(t_1), D^1\psi_k(t_1), \dots, D^{m-1}\psi_k(t_1)]^T &\sim N_m(0, \sigma_{\mu_k}^2 I_m), \\ [B_k(t_1), D^1B_k(t_1), \dots, D^{n-1}B_k(t_1)]^T &\sim N_n(0, \sigma_{\alpha_k}^2 I_n), \end{aligned}$$

while those for the variances in the state equation follow

$$\begin{aligned} \sigma_{\psi_k}^2 &\sim \text{invGa}(a_\psi, b_\psi), \\ \sigma_{B_k}^2 &\sim \text{invGa}(a_B, b_B). \end{aligned}$$

2.4 Hyperparameter interpretation

We now focus our attention on the hyperparameters of the priors for $\sigma_{\xi_{lk}}^2$, $\sigma_{A_k}^2$, $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$. Several simulation studies have shown that the higher the variances in the latent state equations, the better our formulation accommodates locally adaptive smoothing for sudden changes in covariances and means. A theoretical support for this data-driven consideration can be identified in the connection between the nGP prior and nested smoothing splines. It has been shown by Zhu and Dunson (2012) that the posterior mean of U with reference to the problem of nonparametric mean regression under the nGP prior can be related to the minimizer of the equation

$$\frac{1}{T} \sum_{i=1}^T (y_i - U(t_i))^2 + \lambda_U \int_{\mathcal{T}} (D^m U(t) - C(t))^2 dt + \lambda_C \int_{\mathcal{T}} (D^n C(t))^2 dt,$$

where C is the locally instantaneous function and $\lambda_U \in \mathfrak{R}^+$ and $\lambda_C \in \mathfrak{R}^+$ regulate the smoothness of unknown functions U and C respectively, leading to less smoothed patterns when fixed at low values. The resulting inverse relationship between these smoothing parameters and the variances in the state equation, together with the results in the simulation studies, suggest to fix the hyperparameters in the Inverse Gamma prior for $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$, $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$ so as to allow high variances in the case in which the time series analyzed are expected

to have strong changes in their covariance (or mean) dynamic. In practical applications, it may be useful to obtain a first estimate of the covariance matrix $\tilde{\Sigma}(t)$ and the mean vector $\tilde{\mu}(t)$ to set the hyperparameters. More specifically, $\tilde{\mu}_j(t_i)$ can be the output of a standard moving average on each time series $y_j = [y_{j1}, \dots, y_{jT}]$, while $\tilde{\Sigma}(t_i)$ can be obtained by a simple estimator, such as the EWMA procedure. With these choices, the recursive equation

$$\tilde{\Sigma}(t_i) = (1 - \lambda)\{[y_{i-1} - \tilde{\mu}(t_{i-1})][y_{i-1} - \tilde{\mu}(t_{i-1})]^T\} + \lambda\tilde{\Sigma}(t_{i-1})$$

become easy to implement.

3. Posterior Computation

The algorithm for posterior computation alternates between a simple and efficient simulation smoother step (Durbin and Koopman, 2002) to update the state space formulation of the nGP, and standard Gibbs sampling steps for updating the parametric component parameters from their conditional distributions. When also the mean process needs to be estimated, an additional step with a block sampling for $\{\psi(t_i)\}_{i=1}^T$ and $\{\nu_i\}_{i=1}^T$ is implemented.

3.1 Gibbs Sampling

We outline here the main features of the algorithm for posterior computation based on observations (y_i, t_i) for $i = 1, \dots, T$, while the complete algorithm is provided in the Appendix.

- A. Given Θ and $\{\eta_i\}_{i=1}^T$, a multivariate version of the MCMC algorithm proposed by Zhu and Dunson (2012) draws posterior samples from each dictionary element's function $\{\xi_{lk}(t_i)\}_{i=1}^T$, its first order derivative $\{\xi'_{lk}(t_i)\}_{i=1}^T$, the corresponding instantaneous mean $\{A_{lk}(t_i)\}_{i=1}^T$, the variances in the state equations $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$ and the variances of the error terms in the observation equation σ_j^2 with $j = 1, \dots, p$.
- B. If the mean process needs not to be estimated, recalling the prior $\eta_i \sim N_{K^*}(0, I_{K^*})$ and model (1), the standard conjugate posterior distribution from which to sample the vector of latent factors for each i given Θ , $\{\sigma_j^{-2}\}_{j=1}^p$, $\{y_i\}_{i=1}^T$ and $\{\xi(t_i)\}_{i=1}^T$ is Gaussian. Otherwise, if we want to incorporate the mean regression, through model (4), we implement a block sampling of $\{\psi(t_i)\}_{i=1}^T$ and $\{\nu_i\}_{i=1}^T$ following a similar approach used for drawing samples from the dictionary elements process.
- C. Finally, conditioned on $\{y_i\}_{i=1}^T$, $\{\eta_i\}_{i=1}^T$, $\{\sigma_j^{-2}\}_{j=1}^p$ and $\{\xi(t_i)\}_{i=1}^T$, and recalling the shrink-

age prior for the elements of Θ in (8), we update Θ , each local shrinkage hyperparameter ϕ_{jl} and the global shrinkage hyperparameters τ_l following the standard conjugate analysis.

3.2 Online Updating

The problem of online updating represents a key point in multivariate time series with high frequency data. Referring to our formulation, we are interested in updating an approximated posterior distribution for $\Sigma(t_{T+h})$ and $\mu(t_{T+h})$ with $h = 1, \dots, H$ once a new vector of observation $\{y_i\}_{i=T+1}^{T+H}$ is available, instead of rerunning posterior computation for the whole time series.

Using the posterior estimates of the Gibbs sampler based on observations available up to time T , $\{y_i\}_{i=1}^T$, it is easy to implement (see in Appendix) a highly computationally tractable online updating algorithm which alternates between steps *A* and *B* outlined in the previous section for the new set of observations, and that can be initialized at $T + 1$ using the one step ahead predictive distribution for the latent state vector in the state space formulation.

Note that the initialization procedure for latent state vectors in the algorithm depends on the sample moments of the posterior distribution for the latent states at T . As it is known for Kalman smoothers (see, e.g., Durbin and Koopman 2001), this could lead to computational problems in the online updating due to the larger conditional variances of the latent states at the end of the sample (i.e., at T). To overcome this problem, we replace the previous assumptions for the initial values with a data-driven initialization scheme. In particular, instead of using only the new observations for the online updating, we run the algorithm for $\{y_i\}_{i=T-k}^{T+H}$, with k small, and choosing a diffuse but proper prior for the initial states at $T - k$. As a result the distribution of the smoothed states at T is not anymore affected by the problem of large conditional variances leading to better online updating performance.

4. Simulation Studies

The aim of the following simulation study is to assess whether and to what extent the proposed model can accommodate, in practice, even dramatic changes in the time-varying covariances and to compare the performance of our proposal (LBCR, local Bayesian covariance regression) with respect to BCR proposed by Fox and Dunson (2011). In the last subsection we also analyze the performance of the proposed online updating algorithm.

4.1 Estimation Performance

We generate a set of 5-dimensional observations y_i for each t_i in the discrete set $\mathcal{T}_o = \{1, 2, \dots, 100\}$, from the latent factor model in (1) with $\Lambda(t_i) = \Theta\xi(t_i)$ and η_i defined as in (3). To allow dramatic changes of the covariances in the generating mechanism, we consider a 2×2 (i.e. $K = L = 2$) matrix $\{\xi(t_i)\}_{i=1}^{100}$ of time-varying functions adapted from Donoho and Johnstone (1994) with locally-varying smoothness (more specifically we choose ‘bumps’ functions). The latent mean dictionary elements $\{\psi(t_i)\}_{i=1}^{100}$ are simulated from a Gaussian process $GP(0, c)$ with length scale $\kappa = 10$, while the elements in matrix Θ can be obtained from the shrinkage prior in (8) with $a_1 = a_2 = 10$. Finally the elements of the diagonal matrix Σ_0^{-1} are sampled independently from $Ga(1, 0.1)$.

Posterior computation for our proposed approach is performed by using truncation levels $K^* = L^* = 2$; placing a $Ga(1, 0.1)$ prior on the precision parameters σ_j^{-2} and choosing $a_1 = a_2 = 2$. As regards the nGP prior for each dictionary element ξ_{lk} with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$, we choose diffuse but proper priors for the initial values by setting $\sigma_{\mu_{lk}}^2 = \sigma_{\alpha_{lk}}^2 = 100$ and place an $invGa(2, 5 \times 10^8)$ prior on each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ in order to allow less smoothed behavior according to a previous graphical analysis of $\tilde{\Sigma}(t_i)$ estimated via EWMA. Similarly we set $\sigma_{\mu_k}^2 = \sigma_{\alpha_k}^2 = 100$ in the prior for the initial values of the latent state equations resulting from the nGP prior for ψ_k , and consider $a_\psi = a_B = b_\psi = b_B = 0.005$ to balance the rough behavior induced on the nonparametric mean functions by the settings of the nGP prior on ξ_{lk} , as suggested from previous graphical analysis. Note also that for posterior computation, we first scale the predictor space to $(0, 1]$, leading to $\delta_i = 1/100$, for $i = 1, \dots, 100$.

For inference in BCR we consider the same previous hyperparameters setting for Θ and Σ_0 priors as well as the same truncation levels K^* and L^* , while the length scale κ in GP prior for ξ_{lk} and ψ_k has been set to 10 using the data-driven heuristic outlined by Fox and Dunson (2011). In both cases we run 50,000 Gibbs iterations discarding the first 20,000 as burn-in and thinning the chain every 5 samples. Examination of trace plots of the elements of $\{\Sigma(t_i)\}_{i=1}^{100}$ and $\{\mu(t_i)\}_{i=1}^{100}$ showed no evidence against convergence.

Figure 2 compares true and posterior mean (for both approaches) of $\mu(t)$ and $\Sigma(t)$ over the predictor space \mathcal{T}_o together with the point-wise 95% high posterior density intervals. From this plots we can clearly note that our approach is able to capture conditional het-

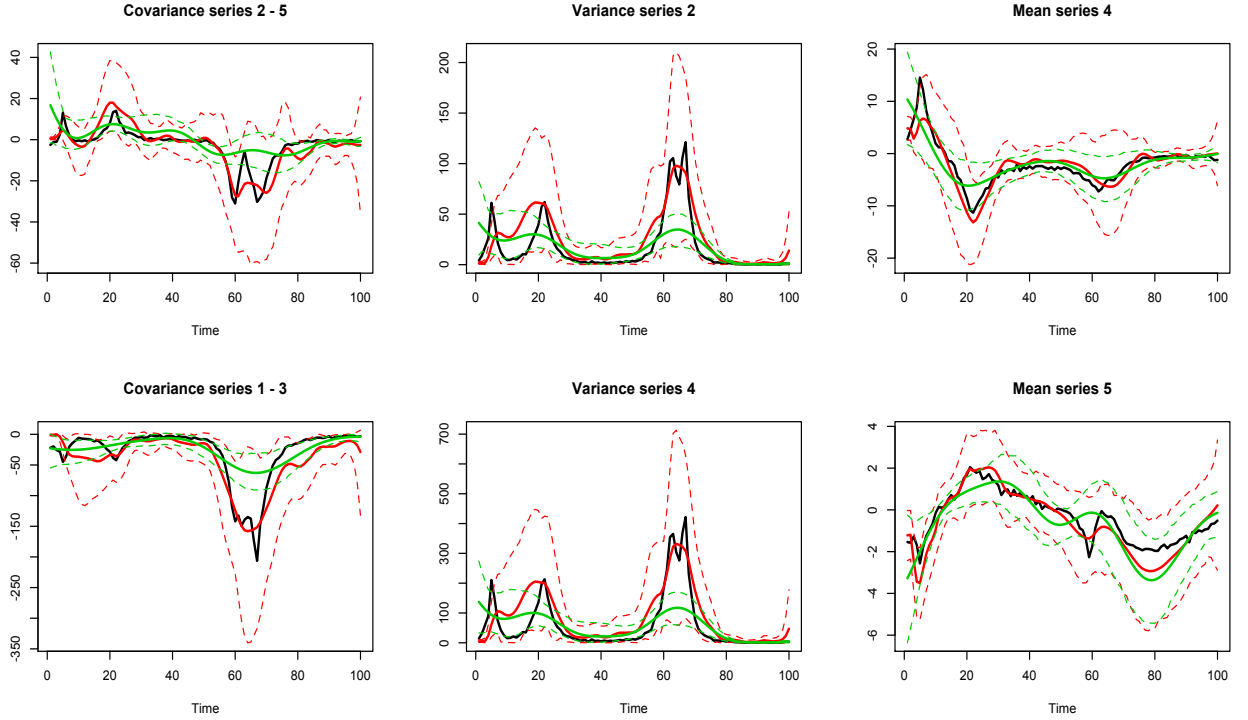


Figure 2: Plots of truth (black) and posterior mean respectively of LBCR (solid red line) and BCR (solid green line) for selected components of the covariance (left), variance (middle), mean (right). For both approaches the dotted lines represent the 95% highest posterior density intervals.

eroscedasticity as well as mean patterns, also in correspondence of dramatic changes in the time-varying true functions. The major differences compared to the true values can be found at the beginning and at the end of the series and are likely to be related to the structure of the simulation smoother which also causes a widening of the credibility bands at the very end of the series, for references regarding this issue see Durbin and Koopman (2001). However, even in the most problematic cases, the true values are within the bands of the 95% highest posterior density intervals. Much more problematic is the behavior of the posterior distributions for BCR which badly over-smooth both covariance and mean functions leading also to many 95% highest posterior density (hpd) intervals not containing the true values. The comparison of the summaries of the squared errors between true values $\{\mu(t_i)\}_{i=1}^{100}$ and $\{\Sigma(t_i)\}_{i=1}^{100}$ and posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{100}$ and $\{\hat{\mu}(t_i)\}_{i=1}^{100}$ respectively for BCR and LBCR in Table 1, once again confirms the overall better performance of our approach.

To better understand the improvement of our approach in allowing locally varying smoothness and to evaluate the consequences of the over-smoothing induced by BCR on

	Mean $\{\mu(t_i)\}$		Covariance $\{\Sigma(t_i)\}$	
	BCR	LBCR	BCR	LBCR
Mean	2.24	1.64	1351.09	667.27
90th Quantile	4.86	3.22	1608.45	1556.38
95th Quantile	9.39	4.98	5140.102	3460.47
Max	74.57	67.87	96092.35	22373.37

Table 1: Summaries of the squared errors between true values $\{\mu(t_i)\}_{i=1}^{100}$ and $\{\Sigma(t_i)\}_{i=1}^{100}$ and posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{100}$ and $\{\hat{\mu}(t_i)\}_{i=1}^{100}$ obtained respectively with BCR and our LBCR.

the distribution of y_i with $i = 1, \dots, 100$ consider Figure 3 which shows, for some selected series $\{y_{ji}\}_{i=1}^{100}$, the time varying mean together with the point-wise 2.5% and 97.5% quantiles of the marginal distribution of y_{ji} induced respectively by the true mean and true variance, the posterior mean of $\mu_j(t_i)$ and $\Sigma_{jj}(t_i)$ from our proposed approach and the posterior mean of the same quantities from the competing alternative. We can clearly see that the marginal distribution of y_{ji} induced by BCR is over-concentrated near the mean leading to incorrect inferences. Note that our proposal is also able to accommodate heavy tails, a typical characteristic in financial series.

4.2 Online Updating Performance

To analyze the performance of the online updating algorithm, we simulate 50 new observations $\{y_i\}_{i=101}^{150}$ with $t_i \in \mathcal{T}_o^* = \{101, \dots, 150\}$, considering the same Θ and Σ_0 used in the generating mechanism for the previous simulated data and taking the 50 subsequent observations of the bumps functions for the dictionary elements $\{\xi(t_i)\}_{i=101}^{150}$; finally the additional latent mean dictionary elements $\{\psi(t_i)\}_{i=101}^{150}$ are simulated as before maintaining the continuity with the previously simulated functions $\{\psi(t_i)\}_{i=1}^{100}$. According to the algorithm described in Section 3.2, we fix Θ , Σ_0 , $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, $\{\sigma_{\psi_k}^2\}$ and $\{\sigma_{B_k}^2\}$ at their posterior mean from the previous Gibbs sampler and consider the last three observations y_{98} , y_{99} and y_{100} (i.e. $k = 3$) to initialize the simulation smoother in $i = 101$ through the proposed data-driven initialization approach. Posterior computation shows good performance in terms of mixing, and convergence is assessed after 5,000 Gibbs iterations with a small burn-in of 500.

Figure 4 compares true mean and covariance to posterior mean of a select set of components of $\{\mu(t_i)\}_{i=101}^{150}$ and $\{\Sigma(t_i)\}_{i=101}^{150}$, including also the 95% hpd intervals. The results clearly show that the online updating is characterized by a good performance which allows

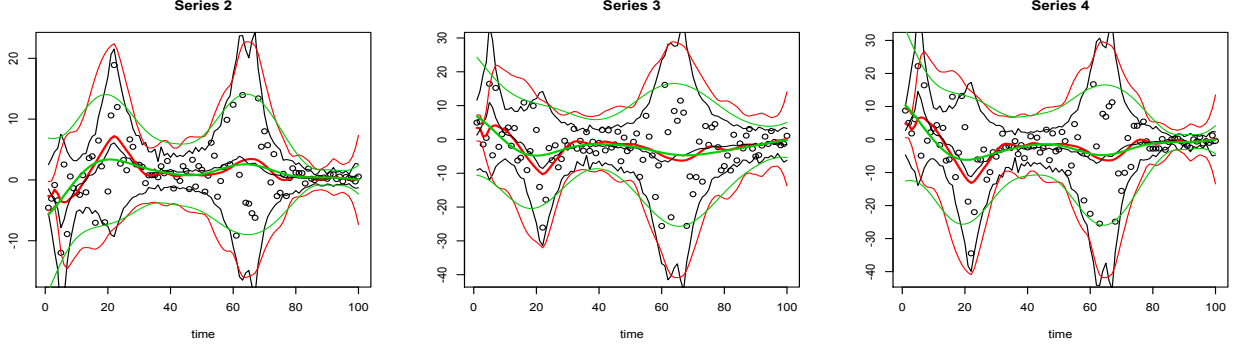


Figure 3: Plot for 3 selected simulated series of the time-varying mean $\mu_j(t_i)$ and the time-varying 2.5% and 97.5% quantiles of the marginal distribution of y_{ji} with true mean and variance (black), mean and variance from posterior mean of LBCR (red), mean and variance from posterior mean of BCR (green). Black points represent the simulated data.

to capture the behavior of new observations conditioning on the previous estimates. Note that the posterior distribution of the approximated mean and covariance functions tends to slightly over-estimate the patterns of the functions at dramatic changes, however also in these cases the true values are within the bands of the credibility intervals. Finally note that the data-driven initialization ensures a good behavior at the beginning of the series, while the results at the end have wide uncertainty bands as expected.

5. Application

Spurred by the recent growth of interest in the dynamic dependence structure between financial markets in different countries, and in its features during the crisis that have followed in recent years, we applied our LBCR to the multivariate time series of the main national stock market indices.

5.1 National Stock Indices (NSI), Introduction and Motivation

National Stock Indices represent technical tools that allow, through the synthesis of numerous data on the evolution of the various stocks, to detect underlying trends in the financial market, with reference to a specific basis of currency and time. More specifically, each Market Index can be defined as a weighted sum of the values of a set of national stocks, whose weighting factors is equal to the ratio of its market capitalization in a specific date and overall of the whole set on the same date.

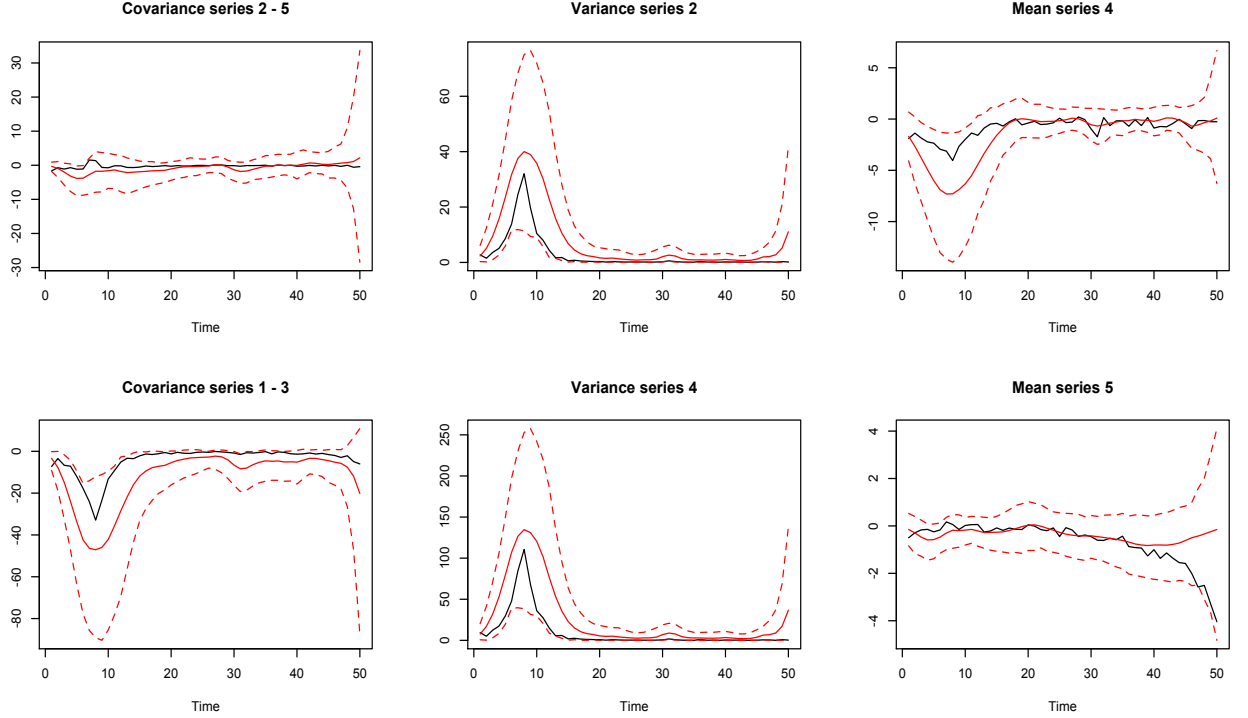


Figure 4: Plots of truth (black) and posterior mean of the online updating procedure (solid red line) for selected components of the covariance (left), variance (middle), mean (right). The dotted lines represent the 95% highest posterior density intervals.

In this application we focus our attention on the multivariate weekly time series of the main 33 (i.e. $p = 33$) National Stock Indices from 12/07/2004 to 25/06/2012 ($T=416$ weeks). Figure 5 shows the main features in terms of stationarity, mean patterns and volatility of two selected NSI downloaded from <http://finance.yahoo.com/>. The non-stationary behavior, together with the different bases of currency and time, motivate the use of logarithmic returns $y_{ji} = \log(I_{ji}/I_{ji-1})$, where I_{ji} is the value of the National Stock Index j at i . Beside this, the marginal distribution of log returns shows heavy tails and irregular cyclical trends in the nonparametric estimation of the mean, while EWMA estimates highlight rapid changes of volatility during the financial crisis observed in the recent years. All these results, together with large settings and high frequency data typical in financial fields, motivate the use of our approach to obtain a better characterization of the time-varying dependence structure among financial markets.

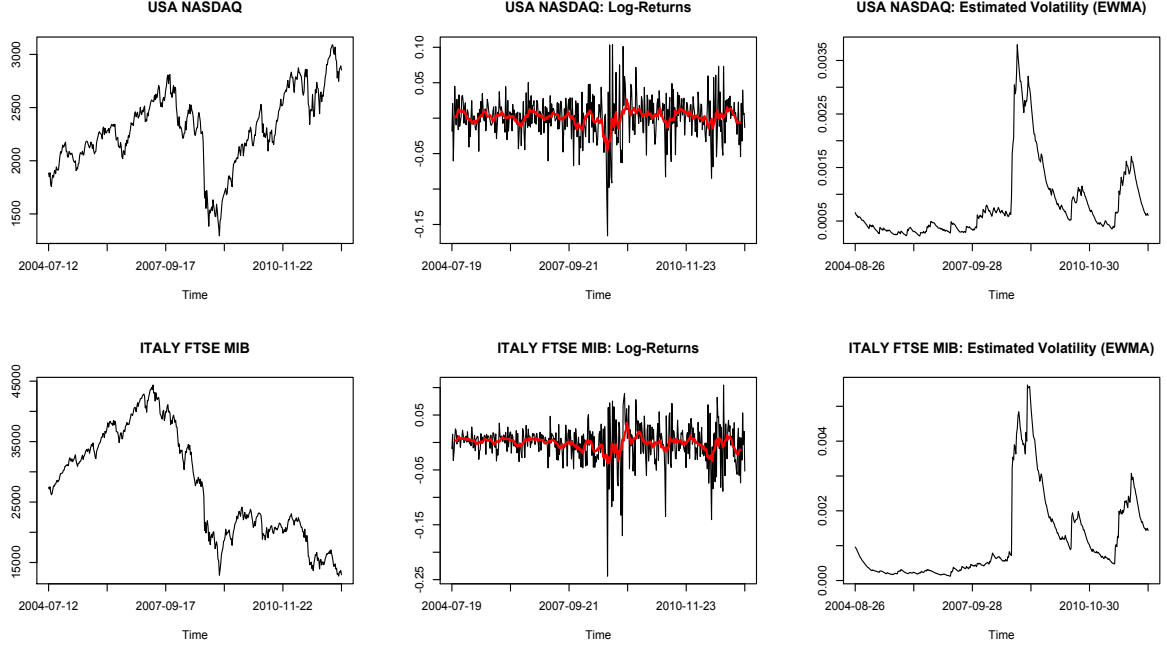


Figure 5: Plots of the main features of USA NASDAQ (top) and ITALY FTSE MIB (bottom). Specifically: observed time series (left), log-returns series (black) with nonparametric mean estimation via 12 week Equally Weighted Moving Average (red) in the middle, EWMA volatility estimates (right).

5.2 LBCR for National Stock Index (NSI)

We consider the heteroscedastic model $y_i \sim N_{33}(\mu(t_i), \Sigma(t_i))$ for $i = 1, \dots, 415$ and t_i in the discrete set $\mathcal{T}_o = \{1, 2, \dots, 415\}$, where mean $\mu(t_i)$ and covariance matrix $\Sigma(t_i)$ of the National Stock Indices at time $t = t_i$ are given in (4) and (2), respectively.

Posterior computation is performed by first rescaling the predictor space \mathcal{T}_o to $(0, 1]$ and using the same setting of the simulation study, with the exception of the truncation levels fixed at $K^* = 4$ and $L^* = 5$ and the hyperparameters of the nGP prior for each ξ_{lk} and ψ_k with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$, set to $a_\xi = a_A = a_\psi = a_B = 2$ and $b_\xi = b_A = b_\psi = b_B = 5 \times 10^7$ to capture also rapid changes in the mean functions according to Figure 5. Missing values in our dataset do not represent a limitation since the Bayesian approach and the use of stochastic processes in continuous time allow us to update our posterior considering solely the observed data. We run 10,000 Gibbs iterations with a burn-in of 2,500. Examination of trace plots for $\{\Sigma(t_i)\}_{i=1}^{415}$ and $\{\mu(t_i)\}_{i=1}^{415}$ showed no evidence against convergence.

Posterior distributions for the variances in Figure 6 demonstrate that we are clearly able to capture the rapid changes in the dynamics of volatility that occur during the world

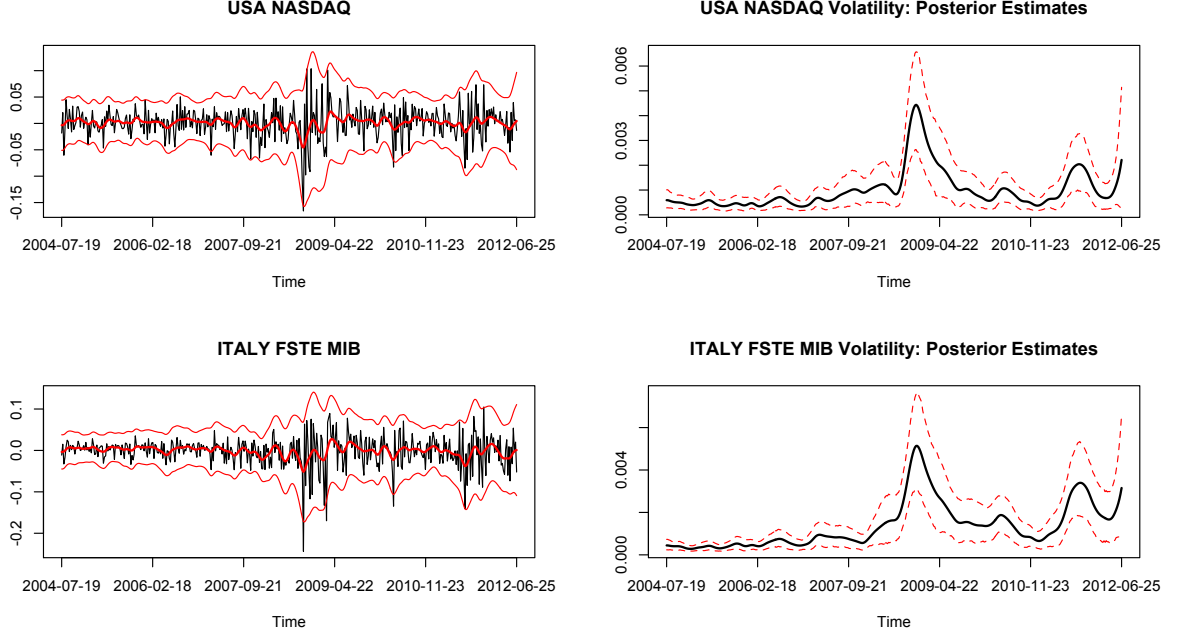


Figure 6: Left: Plot for 2 NSI, respectively USA NASDAQ (top) and ITALY FTSE MIB (bottom), of the log returns (black) and the time-varying estimated mean $\{\hat{\mu}_j(t_i)\}_{i=1}^{415}$ together with the time-varying 2.5% and 97.5% quantiles (red) of the marginal distribution of y_{ji} from LBCR. Right: posterior mean (black) and 95% hpd (dotted red) for the variances $\{\Sigma_{jj}(t_i)\}_{i=1}^{415}$.

financial crisis of 2008, in early 2010 with the Greek debt crisis and in the summer of 2011 with the financial speculation in government bonds of European countries together with the rejection of the U.S. budget and the downgrading of the United States rating. Moreover, the resulting marginal distribution of the log returns induced by the posterior mean of $\mu_j(t)$ and $\Sigma_{jj}(t)$, shows that we are also able to accommodate heavy tails as well as cyclical trends for the means. Important information about the ability of our model to capture the evolution of world geo-economic structure during different finance scenarios are provided in Figures 7 and 8. From the correlations between NASDAQ and the other National Stock Indices (based on the posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{415}$ of the covariances function) in Figure 7, we can immediately notice the presence of a clear geo-economic structure in world financial markets, where the dependence between the U.S. and European countries is systematically higher than that of South East Asian Nations (Economic Tigers), showing also different reactions to crises. Plots at the top of the Figure 8 confirms the above considerations showing how Western countries exhibit more connection with countries closer in terms of geographical, political

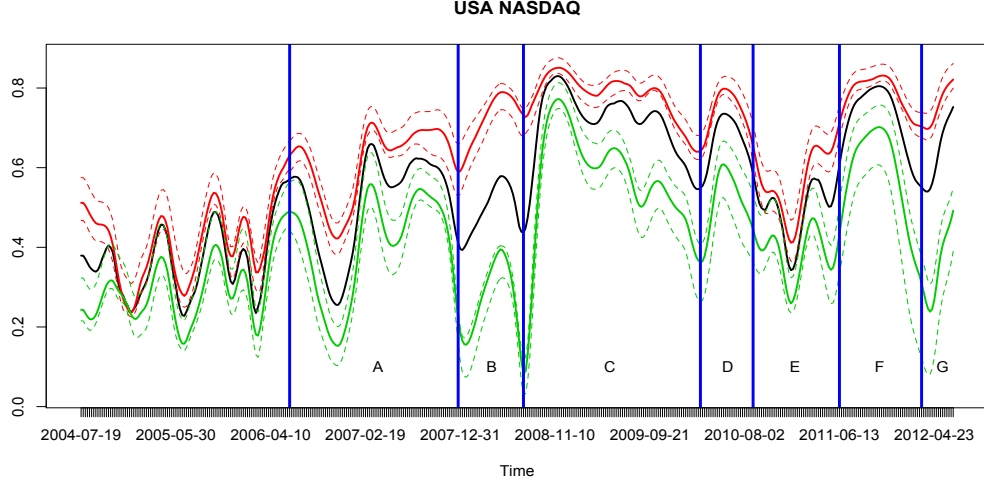


Figure 7: Black line: For USA NASDAQ median of correlations with the other 32 NSI based on posterior mean of $\{\Sigma(t_i)\}_{i=1}^{415}$. Red lines: 25%, 75% (dotted lines) and 50% (solid line) quantiles of correlations between USA NASDAQ and European countries (without considering Greece and Russia which present a specific pattern). Green lines: 25%, 75% (dotted lines) and 50% (solid line) quantiles of correlations between USA NASDAQ and the countries of Southeast Asia (Asian Tigers and India). The timeline is divided in windows that relate to the main financial events of the recent years, specifically event A corresponds to the burst of U.S. housing bubble, event B to the concrete risk of failure of the first U.S. credit agencies (Bear Stearns, Fannie Mae and Freddie Mac), event C to the world financial crisis after the Lehman Brothers' bankruptcy, event D to the Greek debt crisis, event E to financial reform launched by Barack Obama and EU efforts to save Greece (the two peaks represent respectively Irish debt crisis and Portugal debt crisis), event F to the worsening of European sovereign-debt crisis and the rejection of the U.S. budget, finally G to the crisis of credit institutions in Spain and the growing financial instability of the Eurozone.

and economic structure; the same holds for Eastern countries where we observe a reversal of the colored curves. As expected, Russia is placed in a middle path between the two blocks. A further element that our model captures about the structure of the markets is shown in the plots at the bottom of Figure 8. The time-varying regression coefficients obtained from the standard formulas of the conditional normal distribution based on the posterior mean of $\{\mu(t_i)\}_{i=1}^{415}$ and $\{\Sigma(t_i)\}_{i=1}^{415}$ highlight clearly the increasing dependence of European countries with higher crisis in sovereign debt and Germany, which plays a central role in Eurozone as expected.

The flexibility of the proposed approach and the possibility of accommodating varying smoothness in the trajectories over time, allow us to obtain a good characterization of the dynamic dependence structure according with the major theories on financial crisis. Figure 7

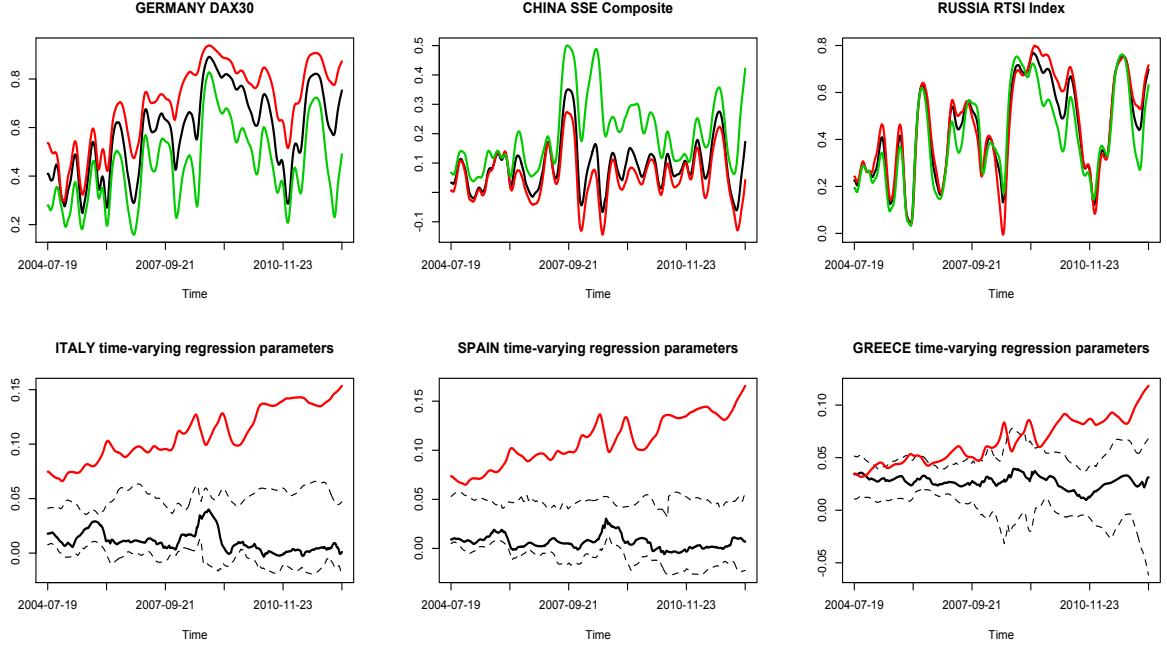


Figure 8: Top: For 3 selected stock market indices, respectively GERMANY DAX30 (top), CHINA SSE Composite (middle) and RUSSIA RTSI Index (bottom), plot of the median of the correlation based on posterior mean of $\{\Sigma(t_i)\}_{i=1}^{415}$ with the other 32 world stock indices (black), the European countries without considering Greece and Russia (red) and the Asian Tigers including India (green). Bottom: For 3 of the European countries more subject to sovereign debt crisis, respectively ITALY (left), SPAIN (middle) and GREECE (right), plot of 25th, 50th and 75th quantiles of the time-varying regression parameters based on posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{415}$ with the other countries (black) and Germany (red).

shows how the change of regime in correlations occurs exactly in correspondence to the burst of the U.S. housing bubble (A), in the second half of 2006. Moreover we can immediately notice that the correlations among financial markets increase significantly during the crises, showing a clear international financial contagion effect in agreement with other theories on financial crisis (see, e.g., Baig and Goldfajn, 1999, Stijn and Forbes, 2009). As expected the persistence of high levels of correlation is evident during the global financial crisis between late-2008 and end-2009 (C), at the beginning of which our approach also captures a dramatic change in the correlations between the U.S. and Economic Tigers, which lead to levels close to those of Europe. Further rapid changes are identified in correspondence of Greek crisis (D), the worsening of European sovereign-debt crisis and the rejection of the U.S. budget (F) and the recent crisis of credit institutions in Spain together with the growing financial instability Eurozone (G). Finally, even in the period of U.S. financial reform launched by

Barack Obama and EU efforts to save Greece (E), we can notice two peaks representing respectively Irish debt crisis and Portugal debt crisis.

5.3 National Stock Indices, Updating and Predicting

The possibility to quickly update the estimates and the predictions as soon as new data arrive, represents a crucial aspect to obtain quantitative information about the future scenarios of the crisis in financial markets. To answer this goal, we apply the online updating algorithm presented in Section 3.2, to the new set of weekly observations $\{y_i\}_{i=416}^{422}$ from 02/07/2012 to 13/08/2012 conditioning on posterior estimates of the Gibbs sampler based on observations $\{y_i\}_{i=1}^{415}$ available up to 25/06/2012. We initialized the simulation smoother algorithm with the last 8 observations of the previous sample.

Plots at the top of Figure 9 show, for 3 selected National Stock Indices, the new observed log returns $\{y_{ji}\}_{i=416}^{422}$ (black) together with the mean and the 2.5% and 97.5% quantiles of the marginal distribution (red) and conditional distribution (green) of $y_{ji}|y_i^{-j}$ with $y_i^{-j} = \{y_{qi}, q \neq j\}$. We use standard formulas of the multivariate normal distribution based on the posterior mean of the updated $\{\Sigma(t_i)\}_{i=416}^{422}$ and $\{\mu(t_i)\}_{i=416}^{422}$ after 5,000 Gibbs iterations with a burn in of 500. This is sufficient for convergence based on examining trace plots of the time-varying mean and covariance matrices. From these results, we can clearly notice the good performance of our proposed online updating algorithm in obtaining a characterization for the distribution of new observations. Also note that the multivariate approach together with a flexible model for the mean and covariance, allow for significant improvements when the conditional distribution of an index given the others are analyzed.

To obtain further information about the predictive performance of our LBCR, we can easily use our online updating algorithm to obtain h step-ahead predictions for $\Sigma(t_{T+h|T})$ and $\mu(t_{T+h|T})$ with $h = 1, \dots, H$. In particular, referring to Durbin and Koopman (2001), we can generate the forecasts $\hat{\Sigma}(t_{T+h|T})$ and $\hat{\mu}(t_{T+h|T})$ for $h = 1, \dots, H$ merely by treating $\{y_i\}_{i=T+1}^{T+H}$ as missing values in the proposed online updating algorithm. Here, we consider the one step ahead prediction (i.e. $H = 1$) problem for the new observations. More specifically, for each i from 415 to 421, we update the mean and covariance functions conditioning on information up to t_i through the online algorithm and then obtain the predicted posterior distribution for $\Sigma(t_{i+1|i})$ and $\mu(t_{i+1|i})$ by adding to the sample considered for the online updating a last

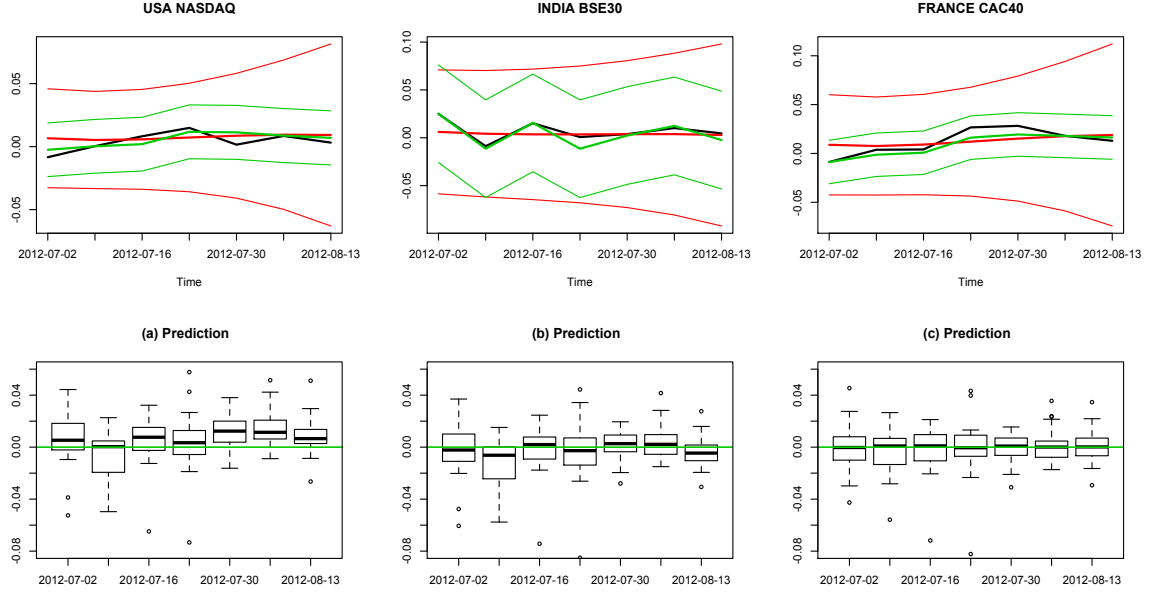


Figure 9: Top: For 3 selected NSI, respectively USA NASDAQ (left), INDIA BSE30 (middle) and FRANCE CAC40 (right), plot of the observed log returns (black) together with the mean and the 2.5% and 97.5% quantiles of the marginal distribution (red) and conditional distribution given the other 32 NSI (green) based on the posterior mean of $\{\Sigma(t_i)\}_{i=416}^{422}$ and $\{\mu(t_i)\}_{i=416}^{422}$ from the online updating procedure for the new observations from 02/07/2012 to 13/08/2012. Bottom: boxplots of the one step ahead prediction errors for the 33 NSI, where the predicted values are respectively: (a) unconditional mean $\{\tilde{y}_{i+1}\}_{i=415}^{421} = 0$, (b) marginal mean of the one step ahead predictive distribution using the online updating procedure for $\{\tilde{y}_{i+1|i}\}_{i=415}^{421}$, (c) conditional mean given the log returns of the other 32 NSI at $i + 1$ of the one step ahead predictive distribution using the online updating procedure for $\{\tilde{y}_{i+1|i}\}_{i=415}^{421}$. Predictions for (b) and (c) are induced by the posterior mean of $\{\Sigma(t_{i+1})\}_{i=415}^{421}$ and $\{\mu(t_{i+1})\}_{i=415}^{421}$ of our LBCR.

column y_{i+1} of missing values.

Plots at the bottom of Figure 9 show the boxplots of the one step ahead prediction errors for the 33 NSI obtained as the difference between the predicted value $\tilde{y}_{j,i+1|i}$ and, once available, the observed log return $y_{j,i+1}$ with $i + 1 = 416, \dots, 422$ corresponding to weeks from 02/07/2012 to 13/08/2012. In (a) we forecast the future log returns with the unconditional mean $\{\tilde{y}_{i+1}\}_{i=415}^{421} = 0$, which is what is often done in practice under the general assumption of zero mean, stationary log returns. In (b) we consider $\tilde{y}_{i+1|i} = \hat{\mu}(t_{i+1|i})$, the posterior mean of the one step ahead predicted nonparametric mean, obtained from the previous proposed approach after 5,000 Gibbs iteration with a burn in of 500. Finally in (c) we suppose that the log returns of all National Stock Indices except that of country j (i.e., $y_{j,i+1}$) become

available at t_{i+1} and, considering $y_{i+1|i} \sim N_p(\hat{\mu}(t_{i+1|i}), \hat{\Sigma}(t_{i+1|i}))$ with $\hat{\mu}(t_{i+1|i})$ and $\hat{\Sigma}(t_{i+1|i})$ posterior mean of the one step ahead predictive distribution respectively for $\mu(t_{i+1|i})$ and $\Sigma(t_{i+1|i})$, we forecast $\tilde{y}_{j,i+1}$ with the conditional mean of $y_{j,i+1}$ given the other log returns at time t_{i+1} .

Comparing boxplots in (a) with those in (b) we can see that our model allows to obtain improvements also in terms of prediction. Furthermore, by analyzing the boxplots in (c) we can notice how our ability to obtain a good characterization of the time-varying covariance structure can play a crucial role also in improving forecasting, since it enters into the standard formula for calculating the conditional mean in the normal distribution.

6. Discussion

In this paper, we have presented a generalization of Bayesian nonparametric covariance regression to obtain a better characterization for mean and covariance temporal dynamics. Maintaining simple conjugate posterior updates and tractable computations in large p settings, our model increases significantly the flexibility of previous approaches as it captures dramatic changes both in mean and covariance dynamics while accommodating heavy tails. Beside these key advantages, the state space formulation enables development of a fast online updating algorithm particularly useful for high frequency data.

The application to the problem of capturing temporal and geo-economic structure between the main financial markets demonstrates the utility of our approach and the improvements that can be obtained in the analysis of multivariate financial time series with reference to (i) heavy tails, (ii) cyclical trends in the mean structure, (iii) dramatic changes in mean and covariance functions, (iii) high dimensional dataset, (iv) online updating with high frequency data (v) missing values and (vi) predictions. Potentially further improvements are possible using a stochastic differential equation model that explicitly incorporates prior information on dynamics.

Appendix

A.1 Posterior Computation in Zero Mean Case

For a fixed truncation level L^* and a latent factor dimension K^* the detailed steps of the Gibbs sampler for posterior computations are:

1. Define the vector of the latent state and the error terms in the state space equation resulting from nGP prior for dictionary elements as

$$\begin{aligned}\Xi_i &= [\xi_{11}(t_i), \xi_{21}(t_i), \dots, \xi_{L^*K^*}(t_i), \xi'_{11}(t_i), \dots, \xi'_{L^*K^*}(t_i), A_{11}(t_i), \dots, A_{L^*K^*}(t_i)]^T \\ \Omega_{i,\xi} &= [\omega_{i,\xi_{11}}, \omega_{i,\xi_{21}}, \dots, \omega_{i,\xi_{L^*K^*}}, \omega_{i,A_{11}}, \omega_{i,A_{21}}, \dots, \omega_{i,A_{L^*K^*}}]^T\end{aligned}$$

Given Θ , $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, Σ_0 and the variances in latent state equations $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$; update $\{\Xi_i\}_{i=1}^T$ by using the simulation smoother in the following state space model

$$y_i = [\eta_i^T \otimes \Theta, 0_{p \times (2 \times L^* \times K^*)}] \Xi_i + \epsilon_i \quad (10)$$

$$\Xi_{i+1} = T_i \Xi_i + R_i \Omega_{i,\xi} \quad (11)$$

Where the observation equation in (10) results by applying the *vec* operator in the latent factor model $y_i = \Theta \xi(t_i) \eta_i + \epsilon_i$. More specifically recalling the property $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$ we obtain

$$\begin{aligned}y_i = \text{vec}(y_i) &= \text{vec}\{\Theta \xi(t_i) \eta_i + \epsilon_i\} \\ &= \text{vec}\{\Theta \xi(t_i) \eta_i\} + \text{vec}(\epsilon_i) \\ &= (\eta_i^T \otimes \Theta) \text{vec}\{\xi(t_i)\} + \epsilon_i.\end{aligned}$$

The state equation in (11) is a joint representation of the equations resulting from the nGP prior on each ξ_{lk} defined in (7). As a result, the $(3 \times L^* \times K^*) \times (3 \times L^* \times K^*)$ matrix T_i together with the $(3 \times L^* \times K^*) \times (2 \times L^* \times K^*)$ matrix R_i reproduce, for each dictionary element the state equation in (7) by fixing to 0 the coefficients relating latent states with different (l, k) (from the independence between the dictionary elements). Finally, recalling the assumptions on $\omega_{i,\xi_{lk}}$ and $\omega_{i,A_{lk}}$, $\Omega_{i,\xi}$ is normally distributed with $E[\Omega_{i,\xi}] = 0$ and $E[\Omega_{i,\xi} \Omega_{i,\xi}^T] = \text{diag}(\sigma_{\xi_{11}}^2 \delta_i, \sigma_{\xi_{21}}^2 \delta_i, \dots, \sigma_{\xi_{L^*K^*}}^2 \delta_i, \sigma_{A_{11}}^2 \delta_i, \sigma_{A_{21}}^2 \delta_i, \dots, \sigma_{A_{L^*K^*}}^2 \delta_i)$.

2. Given $\{\Xi_i\}_{i=1}^T$ sample each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ respectively from

$$\begin{aligned}\sigma_{\xi_{lk}}^2 | \{\Xi_i\} &\sim \text{invGa} \left(a_\xi + \frac{T}{2}, b_\xi + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(\xi'_{lk}(t_{i+1}) - \xi'_{lk}(t_i) - A_{lk}(t_i) \delta_i)^2}{\delta_i} \right) \\ \sigma_{A_{lk}}^2 | \{\Xi_i\} &\sim \text{invGa} \left(a_A + \frac{T}{2}, b_A + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(A_{lk}(t_{i+1}) - A_{lk}(t_i))^2}{\delta_i} \right)\end{aligned}$$

3. Conditioned on Θ , $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, and $\{\xi(t_i)\}_{i=1}^T$ (obtained from Ξ_i), the standard conjugate posterior from which to update σ_j^{-2} is

$$\sigma_j^{-2}|\Theta, \{\eta_i\}, \{y_i\}, \{\xi_{t_i}\} \sim Ga\left(a_\sigma + \frac{T}{2}, b_\sigma + \frac{1}{2} \sum_{i=1}^T (y_{ji} - \theta_j \xi(t_i) \eta_i)^2\right)$$

Where $\theta_{j\cdot} = [\theta_{j1}, \dots, \theta_{jL^*}]$

4. Given Θ , Σ_0 , y_i , and $\xi(t_i)$, the vector of latent factors at each time t can sampled from the Gaussian conditional distribution of $\eta_i|\Theta, \Sigma_0, y_i, \xi(t_i)$

$$N_{K^*}\left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} y_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1}\right)$$

5. Given $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, $\{\xi(t_i)\}_{i=1}^T$ and the hyperparameters ϕ and τ the shrinkage prior on Θ combined with the likelihood for the latent factor model lead to the Gaussian posterior

$$\theta_{j\cdot}|\{\eta_i\}, \{y_i\}, \{\xi(t_i)\}, \phi, \tau \sim N_{L^*}\left(\tilde{\Sigma}_\theta \tilde{\eta}^T \sigma_j^{-2} \begin{bmatrix} y_{j1} \\ \vdots \\ y_{jT} \end{bmatrix}, \tilde{\Sigma}_\theta\right)$$

where $\tilde{\eta}^T = [\xi(t_1)\eta_1, \xi(t_2)\eta_2, \dots, \xi(t_T)\eta_T]$ and

$$\tilde{\Sigma}_\theta^{-1} = \sigma_j^{-2} \tilde{\eta}^T \tilde{\eta} + \text{diag}(\phi_{j1}\tau_1, \dots, \phi_{jL^*}\tau_{L^*})$$

6. The Gamma prior on the local shrinkage hyperparameter ϕ_{jl} implies the standard conjugate posterior given θ_{jl} and τ_l

$$\phi_{jl}|\theta_{jl}, \tau_l \sim Ga\left(2, \frac{3 + \tau_l \theta_{jl}^2}{2}\right)$$

7. Conditioned on Θ and τ , sample the global shrinkage hyperparameters from

$$\begin{aligned} \vartheta_1|\Theta, \tau^{(-1)} &\sim Ga\left(a_1 + \frac{pL^*}{2}, 1 + \frac{1}{2} \sum_{l=1}^{L^*} \tau_l^{(-1)} \sum_{j=1}^p \phi_{jl} \theta_{jl}^2\right) \\ \vartheta_h|\Theta, \tau^{(-h)} &\sim Ga\left(a_2 + \frac{p(L^* - h + 1)}{2}, 1 + \frac{1}{2} \sum_{l=1}^{L^*} \tau_l^{(-h)} \sum_{j=1}^p \phi_{jl} \theta_{jl}^2\right) \end{aligned}$$

Where $\tau_l^{(-h)} = \prod_{t=1, t \neq h}^l \vartheta_t$ for $h = 1, \dots, p$

A.2 Incorporating unknown mean

The Gibbs sampler for the more general model follows exactly the steps previously outlined except for the step 4 which is replaced by a block sampling of $\{\psi(t_i)\}_{i=1}^T$ and $\{\nu_i\}_{i=1}^T$. In particular

4.1. Similarly to Ξ_i and $\Omega_{i,\xi}$ let

$$\begin{aligned}\Psi_i &= [\psi_1(t_i), \psi_2(t_i), \dots, \psi_{K^*}(t_i), \psi'_1(t_i), \dots, \psi'_{K^*}(t_i), B_1(t_i), \dots, B_{K^*}(t_i)]^T \\ \Omega_{i,\psi} &= [\omega_{i,\psi_1}, \omega_{i,\psi_2}, \dots, \omega_{i,\psi_{K^*}}, \omega_{i,B_1}, \omega_{i,B_2}, \dots, \omega_{i,B_{K^*}}]^T\end{aligned}$$

be the vectors of the latent state and error terms in the state space equation resulting from nGP prior for ψ .

Conditional on Θ , $\{\xi(t_i)\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, Σ_0 , and the variances in latent state equations $\{\sigma_{\psi_k}^2\}$, $\{\sigma_{B_k}^2\}$, with $k = 1, \dots, K^*$; sample $\{\Psi_i\}_{i=1}^T$ from the simulation smoother in the following state space model

$$y_i = [\Theta\xi(t_i), 0_{p \times (2 \times K^*)}] \Psi_i + \varpi_i, \quad (12)$$

$$\Psi_{i+1} = G_i \Psi_i + F_i \Omega_{i,\psi}, \quad (13)$$

$\varpi_i \sim N(0, \Theta\xi(t_i)\xi(t_i)^T\Theta^T + \Sigma_0)$. The observation equation in (12) results by marginalizing out ν_i in the latent factor model with nonparametric mean regression $y_i = \Theta\xi(t_i)\psi(t_i) + \Theta\xi(t_i)\nu_i + \epsilon_i$. Analogously to Ξ_i , the state equation in (13) is a joint representation of the state equation induced by the nGP prior on each ψ_k defined in (9); where the $(3 \times K^*) \times (3 \times K^*)$ matrix G_i and the $(3 \times K^*) \times (2 \times K^*)$ matrix F_i are constructed with the same goal of the matrices T_i and R_i in the state space model for Ξ_i . Finally, $\Omega_{i,\psi} \sim N_{2 \times K^*}(0, \text{diag}(\sigma_{\psi_1}^2 \delta_i, \sigma_{\psi_2}^2 \delta_i, \dots, \sigma_{\psi_{K^*}}^2 \delta_i, \sigma_{B_1}^2 \delta_i, \sigma_{B_2}^2 \delta_i, \dots, \sigma_{B_{K^*}}^2 \delta_i))$.

4.2. Given $\{\Psi_i\}_{i=1}^T$ update each $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$ respectively from

$$\begin{aligned}\sigma_{\psi_k}^2 | \{\Psi_i\} &\sim \text{invGa} \left(a_\psi + \frac{T}{2}, b_\psi + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(\psi'_k(t_{i+1}) - \psi'_k(t_i) - B_k(t_i)\delta_i)^2}{\delta_i} \right) \\ \sigma_{B_k}^2 | \{\Psi_i\} &\sim \text{invGa} \left(a_B + \frac{T}{2}, b_B + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(B_k(t_{i+1}) - B_k(t_i))^2}{\delta_i} \right)\end{aligned}$$

4.3. Conditioned on Θ , Σ_0 , y_i , $\xi(t_i)$ and $\psi(t_i)$, and recalling $\nu_i \sim N_{K^*}(0, I_{K^*})$; the standard conjugate posterior distribution $\nu_i | \Theta, \Sigma_0, \tilde{y}_i, \xi(t_i), \psi(t_i)$ is

$$N_{K^*} \left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} \tilde{y}_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \right)$$

with $\tilde{y}_i = y_i - \Theta\xi(t_i)\psi(t_i) = \Theta\xi(t_i)\nu_i + \epsilon_i$.

B Online Updating Algorithm

Consider Θ , Σ_0 , $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, $\{\sigma_{\psi_k}^2\}$ and $\{\sigma_{B_k}^2\}$ fixed at their posterior mean $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{\hat{\sigma}_{\xi_{lk}}^2\}$, $\{\hat{\sigma}_{A_{lk}}^2\}$, $\{\hat{\sigma}_{\psi_k}^2\}$, $\{\hat{\sigma}_{B_k}^2\}$ respectively, and let $\hat{\Xi}_T$, $\hat{\Sigma}_{\Xi_T}$ and $\hat{\Psi}_T$, $\hat{\Sigma}_{\Psi_T}$ be the sample mean and covariance matrix of the posterior distribution respectively for Ξ_T and Ψ_T obtained from the posterior estimates of the Gibbs sampler conditioned on $\{y_i\}_{i=1}^T$.

1. Given $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{\hat{\sigma}_{\xi_{lk}}^2\}$, $\{\hat{\sigma}_{A_{lk}}^2\}$, $\{\eta_i\}_{i=T+1}^{T+H}$ and $\{y_i\}_{i=T+1}^{T+H}$ update $\{\Xi_i\}_{i=T+1}^{T+H}$ by using the simulation smoother in the following state space model

$$\begin{aligned} y_i &= [\eta_i^T \otimes \hat{\Theta}, 0_{p \times (2 \times K^* \times L^*)}] \Xi_i + \epsilon_i \\ \Xi_{i+1} &= T_i \Xi_i + R_i \Omega_{i,\xi} \end{aligned}$$

Where Ξ_{T+1} can be initialized from the standard one step ahead predictive distribution for the state space model $\Xi_{T+1} \sim N(T_T \hat{\Xi}_T, T_T \hat{\Sigma}_{\Xi_T} T_T^T + R_T E[\Omega_{T,\xi} \Omega_{T,\xi}^T] R_T^T)$

2. Conditioned on $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{\hat{\sigma}_{\psi_k}^2\}$, $\{\hat{\sigma}_{B_k}^2\}$, $\{\xi(t_i)\}_{i=T+1}^{T+H}$ and $\{y_i\}_{i=T+1}^{T+H}$ sample $\{\Psi_i\}_{i=T+1}^{T+H}$ through the simulation smoother in the state space model

$$\begin{aligned} y_i &= [\hat{\Theta}\xi(t_i), 0_{p \times (2 \times K^*)}] \Psi_i + \varpi_i \\ \Psi_{i+1} &= G_i \Psi_i + F_i \Omega_{i,\psi} \end{aligned}$$

Similarly to Ξ_{T+1} , $\Psi_{T+1} \sim N(G_T \hat{\Psi}_T, G_T \hat{\Sigma}_{\Psi_T} G_T^T + F_T E[\Omega_{T,\psi} \Omega_{T,\psi}^T] F_T^T)$

3. Given $\hat{\Theta}$, $\hat{\Sigma}_0$, $\{y_i\}$, $\xi(t_i)$ and $\psi(t_i)$, for $i = T+1, \dots, T+H$, sample ν_i from the standard conjugate posterior distribution for $\nu_i | \Theta, \Sigma_0, \tilde{y}_i, \xi(t_i), \psi(t_i)$:

$$N_{K^*} \left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} \tilde{y}_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \right)$$

with $\tilde{y}_i = y_i - \Theta\xi(t_i)\psi(t_i) = \Theta\xi(t_i)\nu_i + \epsilon_i$.

4. Compute the updated covariance $\{\Sigma(t_i)\}_{i=T+1}^{T+H}$ and mean $\{\mu(t_i)\}_{i=T+1}^{T+H}$ from the usual equations

$$\begin{aligned} \Sigma(t_i) &= \hat{\Theta}\xi(t_i)\xi(t_i)^T \hat{\Theta}^T + \hat{\Sigma}_0 \\ \mu(t_i) &= \hat{\Theta}\xi(t_i)\psi(t_i) \end{aligned}$$

Acknowledgemnts

This research was partially supported by grant R01ES17240 from the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH) and by grant CPDA097208/09 from the University of Padua, Italy.

References

- Aguilar O. and West M. (2000). "Bayesian dynamic factor models and portfolio allocation." *Journal of Business & Economic Statistics*, 18, 338-357.
- Alexander, C.O. (2001). "Orthogonal GARCH." *Mastering Risk*, 2, 21-38.
- Bhattacharya, A. and Dunson, D.B. (2011). "Sparse Bayesian infinite factor models." *Biometrika*, 98, 291-306.
- Bollerslev, T., Engle, R.F. and Wooldrige, J.M. (1988). "A capital-asset pricing model with time-varying covariances." *Journal of Political Economy*, 96, 116-131.
- Burns, P. (2005). "Multivariate GARCH with Only Univariate Estimation." <http://www.burns-stat.com>.
- Carvalho, C.M., Lucas, J.E., Wang, Q., Chang, J., Nevins, J.R. and West M. (2008). "High-dimensional sparse factor modelling - Applications in gene expression genomics." *Journal of the American Statistical Association*, 103, 1438-1456.
- Ding, Z. (1994). "Time series analysis of speculative returns." PhD thesis, University of California, San Diego.
- Donoho, D.L. and Johnstone, J.M. (1994). "Ideal spatial adaptation by wavelet shrinkage." *Biometrika*, 81, 425-455.
- Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press Inc., New York.
- Durbin, J. and Koopman, S. (2002). "A simple and efficient simulation smoother for state space time series analysis." *Biometrika*, 89, 603-616.

- Engle, R.F. and Krones, K.F. (1995). "Multivariate simultaneous generalized ARCH." *Econometric Theory*, 11, 122-150.
- Fox E. and Dunson D.B. (2011). "Bayesian Nonparametric Covariance Regression." *arXiv:1101.2017*.
- Geweke J. and Zhou G. (1996). "Measuring the pricing error of the arbitrage pricing theory." *Review of Financial Studies*, 9, 557-587.
- Harvey, A., Ruiz, E. and Shephard, N. (1994). "Multivariate Stochastic Variance Models." *Review of Economic Studies*, 61, 247-264.
- Lopes, H.F. and West M. (2004). "Bayesian model assessment in factor analysis." *Statistica Sinica*, 14, 41-68.
- Lopes, H. F., Salazar, E. and Gamerman, D. (2008). "Spatial Dynamic Factor Analysis." *Bayesian Analysis*, 3, 759-792.
- Lopes, H.F., Gamerman, D. and Salazar, E. (2011). "Generalized spatial dynamic factor models.", *Computational Statistics & Data Analysis*, 55, 1319-1330.
- Nakajima, J. and West, M. (2012). "Dynamic factor volatility modeling: A Bayesian latent threshold approach", *Journal of Financial Econometrics*, in press.
- Pati D., Bhattacharya A., Pillai N.S. and Dunson D.B. (2012). "Bayesian high-dimensional covariance matrix estimation." <http://ftp.stat.duke.edu/WorkingPapers/12-05.html>.
- Philipov A. and Glickman, M.E. (2006a). "Multivariate stochastic volatility via Wishart processes." *Journal of Business & Economic Statistics*, 24, 313-328.
- Prado R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*. Chapman & Hall / CRC, Boca Raton, FL.
- Ruiz, E. (1994). "Quasi-maximum likelihood estimation of stochastic volatility models." *Journal of Econometrics*, 63, 289-306.
- Tsay, R.S. (2005). *Analysis of Financial Time Series*. Wiley.

- van der Weide, R. (2002), “GO-GARCH: a multivariate generalized orthogonal GARCH model.” *Journal of Applied Econometrics*, 17, 549-564.
- West M. (2003). “Bayesian factor regression models in the large p , small n paradigm.” *Bayesian Statistics*, 7, 723-732.
- Zhu B. and Dunson D.B., (2012). “Locally Adaptive Bayes Nonparametric Regression via Nested Gaussian Processes.” *arXiv:1201.4403*.